Constraint-based approach in geological map generalization

Sayidov, Azimjon ; Weibel, Robert

Abstract: Given the importance of natural resources exploration, generalization in geological mapping is receiving increasing interest, though only few reliable automated generalization tools are available for this purpose today. Thus, improvements to methods for the generalization of categorical data, such as geological or soil maps are in demand. In this short paper, we advocate a constraint-based approach for geological map generalization, which could be implemented by integrating vector and raster based generalization methods. We start by laying out the conceptual framework of this integrated approach. We then continue by defining some sample constraints that can be used to control the generalization process. And we show some initial implementations and experiments with generalization operations in the vector and raster domain.

Originally published at:
Sayidov, Azimjon; Weibel, Robert (2016). Constraint-based approach in geological map generalization. In: 19th ICA Workshop on Generalisation and Multiple Representation, Helsinki (Finnland), 14 June 2016. s.n., online.

Constraint-based approach in geological map generalization

Azimjon Sayidov and Robert Weibel
Department of Geography, University of Zurich
Winterthurerstrasse 190, CH-8057 Zurich
{azimjon.sayidov | robert.weibel}@geo.uzh.ch

Abstract

Given the importance of natural resources exploration, generalization in geological mapping is receiving increasing interest, though only few reliable automated generalization tools are available for this purpose today. Thus, improvements to methods for the generalization of categorical data, such as geological or soil maps are in demand. In this short paper, we advocate a constraint-based approach for geological map generalization, which could be implemented by integrating vector and raster based generalization methods. We start by laying out the conceptual framework of this integrated approach. We then continue by defining some sample constraints that can be used to control the generalization process. And we show some initial implementations and experiments with generalization operations in the vector and raster domain.

1. Introduction

While most research in map generalization has focused on point, line and small-area features typically found on topographic maps, the generalization of categorical maps has received less attention. However, given the interest in natural resources exploration, categorical maps, such as geological or soil maps, are moving into focus again. Categorical maps may either be represented as polygonal subdivisions of vector data or color-coded pixels of raster data. For the analysis and processing of both vector and raster data, many functions exist in commercial or open source systems that have the potential of being used as building blocks of categorical map generalization. Hence, simple polygon processing tools and image analysis operations have been successfully used in parts of the geological map production process. Yet, specific tools and integrated workflows for categorical data generalization are still lacking.

Early research aiming at generalization in a raster environment was carried out by Schylberg (1993) or Su et al. (1997). In vector representations Muller and Wang (1992), Downs and Mackaness (2002), Steiniger and Weibel (2005) and McCabe (2008) provide examples. The integration of both methods was addressed by Peter and Weibel (1999) and Smirnoff et al. (2008). Since both the vector and raster representations have their strengths and weaknesses, an integrated approach exploiting the advantages of both worlds seems particularly attractive. Thus, the conceptual approach used in this paper is based on earlier work by Weibel and Dutton (1998) and Peter and Weibel (1999) and aims to improve existing methods for the generalization of geological maps by integrating vector and raster approaches and keeping control of the quality on the basis of previously defined constraints.

2. Geological maps and their peculiarities

The objective of a geological map is “to interpretively portray the spatial and temporal relationships of rocks, unconsolidated earth materials, and landforms at the earth’s surface” (Jirsa & Boerboom 2003: 2).

Geological maps are among the most complex thematic maps, with various elaborate shapes and structures, rendering the generalization process more complicated and requiring in-depth analysis of these structures prior to the generalization process.

Figure 1 shows map examples of increasing complexity. In the simple map extract of Figure 1A only few geological units are involved, with relatively simple shapes, which could be generalized with the simplification operators. The next level of complexity is many small polygons of the same or similar geological unit (Fig. 1B), which may be generalized by aggregating units or subunits to a unique unit. Another level consists of series of elongated polygons of the same units embedded in, and possibly crossing, other units (Fig. 1C), where the purpose would be to merge neighboring units but try to maintain their overall arrangement (typification). Another complex type is tree-like, dendritic shapes which were formed in the later stage of the quaternary period by river streams carrying along sediments and other minerals and which also describes the position of a river (Fig. 1D). Thus, while simplifying the tree branches topological consistency based on the map scale should be considered. Here, several tree branches could be typified and replaced by a smaller number of simplified and aggregated tree branches. Different types of units, small and big, long and narrow and tree-like features may be gathered in a small space making the process even more complex (Fig. 1E). Generalizing such complex fabrics requires making multiple, possibly interrelated (and conflicting) generalization decisions. Such situations can best be formalized and controlled by using constraints.

The next section gives a brief review of different approaches published in the domain of geological map generalization, which have the potential to cope with the above peculiarities of geological maps.

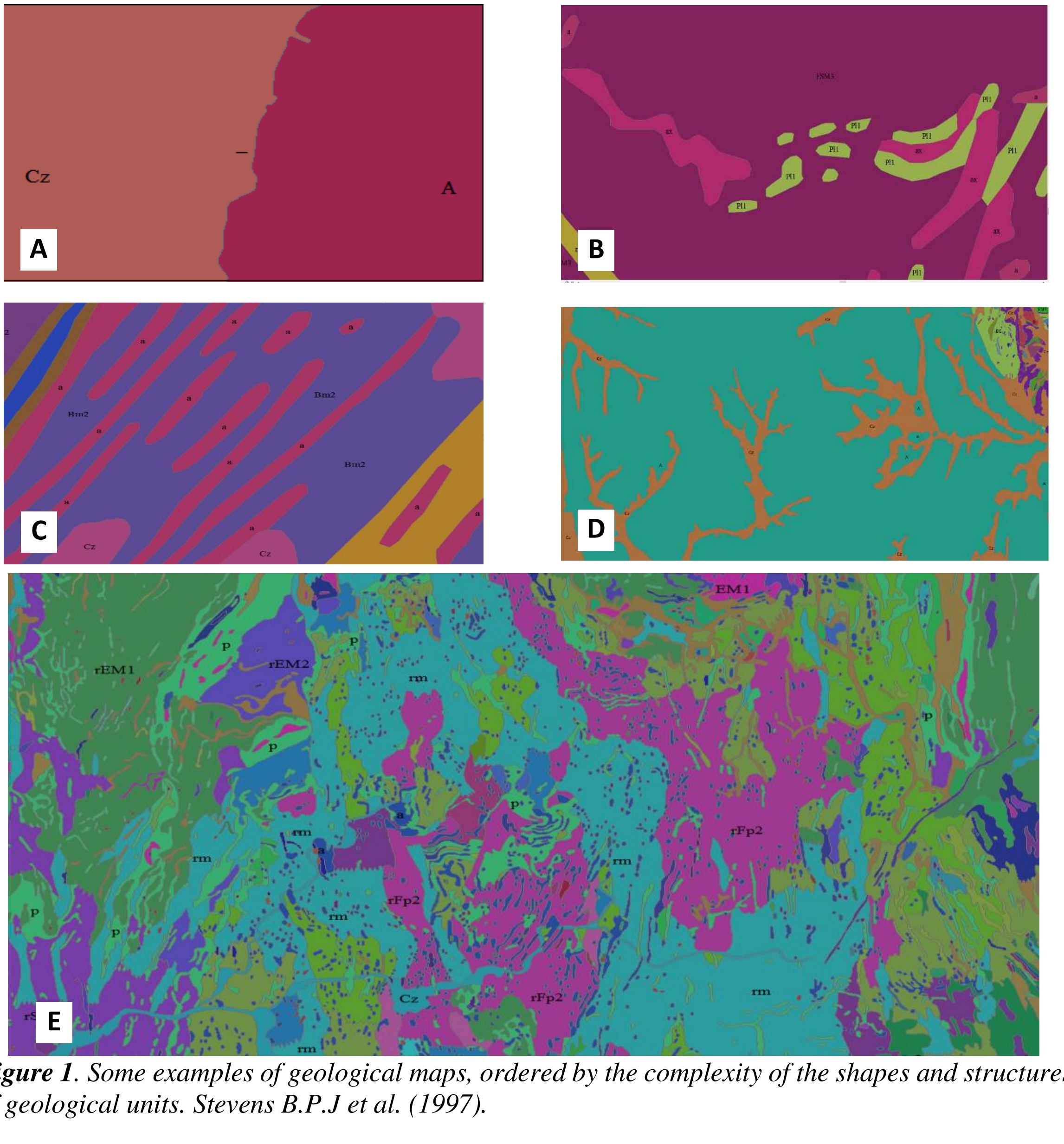

***Figure 1**. Some examples of geological maps, ordered by the complexity of the shapes and structures of geological units. Stevens B.P.J et al. (1997).*

3. Related work

Before moving to the review of the literature relevant for geological map generalization, it is important to point out the necessity of improving the existing generalization packages. Several years ago, this observation was made in a EuroSDR project, which concluded that none of the tested software systems for automated generalization achieved good results for automating the overall generalization process, highlighting least flexibility in the customization of the process (Stoter et al., 2010), which would be crucial for the generalization of complex map types such as geological maps. The situation has not improved a great deal since then.

The earliest noticeable effort in geological map generalization was by Downs and Mackaness (2002), attempting to automate the generalization process of producing a 1:250 000 geological map from a 1:50 000 source map, which is a product of the British Geological Survey (BGS), adopting the conceptual model earlier proposed by Brassel and Weibel (1998). The BGS specialists back then assessed the results as satisfactory and expressed an interest in further research on the topic. However, it was also concluded that the approach is dependent on the specific geological map and requires human operator intervention, making it less flexible and more subjective.

The conceptual workflow model developed by Steiniger and Weibel (2005) is dedicated to the generalization of thematic maps and has three major stages: structural analysis, generalization and visualization. Of these, the structural analysis (termed "structure recognition" in the framework by Brassel and Weibel 1998) is particularly important, since once all the structures are recognized, it becomes possible to decide "when to generalize" and "how to generalize" (Shea and McMaster 1989). The second step consists of constraint-based modeling in a multi-agent system, which makes the concept more objective and flexible. Though keeping the human in the loop ensures the purpose and better communication of a map, it may also hinder objectivity of the overall process.

Inspired by the conceptual model of Steiniger and Weibel (2005), a generalization workflow based on ArcGIS tools was developed by McCabe (2008), thus also pointing out the limitations of the ArcGIS generalization tools. A geological map of the scale 1:24 000 of Santa Rosa, California was generalized to three scales, 1:50 000, 1:100 000 and 1:250 000. Results were compared to the geological maps produced by the US Geological Survey (USGS). The work could be summarized as a generalization experiment, applying the limited tools of ArcGIS available at the time. Nevertheless, the author concluded that the results obtained from the automated tool were better legible and thus encouraging.

Another experiment on the efficiency ArcGIS was carried out by Smirnoff et al. (2008), also supported by the later publication of Smirnoff et al. (2012), comparing their own cellular automata (CA) principle for geological map generalization, which concludes that the cell-based model or cellular automata has important advantages for the automated generalization of geological maps. In the latter research a plug-in called 'GeoScaler' to ArcGIS was developed and tested for surficial and bedrock maps. The results were evaluated and found adequate, while obtaining repeatable results by keeping some amount of human intervention in the process.

4. Methodology

Our conceptual framework adopts the steps of the approach suggested by Peter and Weibel (1999): setting the map controls, defining constraints, defining measures, modeling the process and finally executing the process, while monitoring quality evaluation. Moreover, it may also be regarded as a dynamic generalization model guided by constraints, where decisions depend on the semantic and geometrical characteristics of an object or set of objects, requiring existence of procedural knowledge in order to appropriately select operators and algorithms (Ruas and Plazanet 1996).

One of the fundamental elements of cartographic generalization are the *map controls* (Brassel and Weibel 1988), especially the purpose and scale of the map, which have a direct effect on the content of the map, but also the spatial data model that is used to represent reality. In categorical maps typically the entire surface of the map is covered with continuous polygons or areal features, with no holes nor overlaps. Such maps can equally be modeled as vector or raster data model.

Raster generalization is seen by some authors as a preferred choice and ideal for geological mapping at all scales (Marjoribanks 2010), using classification, reclassification, majority filters, or low and high pass filters. Additionally, various morphological filters, such as erosion, dilation, opening, or closing could be used with different filter elements; as well as least-cost operations (Peter & Weibel 1999).

Conversely, the vector representation lends itself better to geometrical transformations of vertices, such as shifting the position of individual vertices, or removing vertices altogether. Also, since geological units are modeled as entire polygons rather than simply a collection of pixels, spatial relations between polygons can be explicitly modeled, enabling better contextual operations, such as contextual aggregation of sub-categories to a unique category. For instance, in the stratigraphic hierarchy unit types are merged to a higher hierarchy level, such as Dm (1) and Db (2) and eventually merged to D (3) based on the scale of the output map. This process, however, should not only be controlled by attribute values, but also by spatial relations, such as spatial proximity.

$$Dmc + Dmf + Dmj + Dmt = Dm\ (1),$$

$$Dbh + Dbk + Dbm + Dbp + Dbs + Dby = Db\ (2),$$

$$Dm + Db = D\ (3).$$

The next main steps of the framework of Peter and Weibel (1999) consists in defining the generalization *constraints*, and in defining the measures that can implement the previously defined constraints and thus assess whether any constraints are violated. Examples of these steps are given in the next section. It should be noted that constraints and associated measures also play a crucial role in structure recognition (Brassel and Weibel 1988).

5. Constraints and generalization of geological maps

From the perspective of map generalization constraints can be defined as a design specification to which the solutions to a generalization problem should adhere (Weibel and Dutton 1998). Constraints dictate the decisions, limit the search space of the process and restrict the content of the map, while generalizing it. They apply at every stage of the map production process, incl. map design, development of the content, composing the legend of the map and in the portrayal of a map.

Although constraints can be defined regardless of the spatial data model used, vector or raster, their *implementation* may differ. For instance, if the pixel size of a raster is already larger than the minimum visual separation limit, the associated constraints (minimum size, minimum separation distance) will not apply. Similarly, the *measures* used to implement the constraints will differ between the two spatial data models. For instance, distances are measured differently in vector or raster data.

In the process of generalization constraints have following functions:

- conflict detection: to identify areas that have to be generalized, for example by evaluating the quantity and severity of constraint violations
- conflict resolution: to guide the choice of operators according to constraints priorities
- quality evaluation: to control the effect of an algorithm by detecting constraint violations on objects after each transformation (Ruas & Plazanet 1996)

Classifications of the constraints are well described in Weibel and Dutton (1998), who propose:

Graphical constraints, which are related to readability of the map features, such as size, width and differentiation of the objects and which are detected by graphical limits (Weibel and Dutton 1998). Graphical constraints may be considered from two perspectives: within *individual* objects (minimum area, internal width etc. of each area object separately) vs. constraints that are *contextual* and apply to multiple objects, such as the minimum separability distance *between* features.

Regarding so-called "minimum dimensions", that is, the visual limits used, different researchers and organizations report different figures, often also influenced by the type of map, the purpose, and the legend of the map (i.e. the map controls). For instance, it is common practice to use one vertex per mm or maximum two vertices per mm as a minimum differentiation unit between consecutive vertices (Lepland 2014), translating to one vertex in 50 meters at 1:50 000 scale. McCabe (2008) advises that 7000 m^2 is the minimum area at the scale of 1:50 000, whereas Spiess at al. (2002) state 0.80 mm as minimum size for a colored area with a contour, which equals to 1600 m^2 and is almost five times smaller than McCabe's figure. Spiess et al. also provide a series of other useful minimum dimensions.

Topological constraints are related to geometrical accuracy and may accrue after operators such as elimination, aggregation, or simplification have been applied; or if data has been converted between vector and raster data types; or if the topological type was changed, such as in a polygon-to-point collapse, a line-to-point collapse, or a multipoint-to-polygon aggregation. Topological constraints are meant to ensure the topological relationships of connectivity, adjacency and containment. Moreover, self-intersection and overlapping of polygons are not allowed, and this part could be controlled by simple rules included in generalization algorithms, restricting these violations.

Structural constraints define criteria that represent the spatial and semantic structure of the data, where the former handle preserving shapes and patterns that are typical, whereas the semantic structural constraints preserve the semantic logical relationship when features are aggregated, such as rock units merged based on their attribute of the geological unit hierarchy.

This short paper only deals with the initial stage of the generalization of geological maps, with the constraints only related to individual or few polygons, with selected examples presented below:

Minimum size: polygons should not be smaller than the minimum size limit (which may vary according to author; see above); Figure 2A. If violated, this constraint can be enforced by an elimination operation or, if the polygon is considered important (e.g. due to uniqueness of the geological unit), by enlargement.

Minimum distance: Distance between consecutive vertices must not be less than the readability unit. Thus, vertices falling below this threshold should be removed by line simplification. Likewise, collinear vertices can be removed in any case, as they do not modify the shape (Fig. 2B).

***Interior width*:** The interior width of polygons should not be less than minimum separability unit (Fig. 2C). To remedy conflicts with this constraint, the polygon can be locally enlarged.

***Polygon separability*:** The distance between two polygons should not be less than minimum readability unit (Fig. 2D). Otherwise, different options exist: aggregation of the polygons, local enlargement of the isthmus, or displacement.

Self-intersection: Polygon outlines should not intersect with themselves. Avoiding self-intersection is approached by using topology rules (e.g. in ArcGIS geodatabase 'polygons should not intersect').

***Aggregation*:** Groups of disjoint polygons of the same class may be aggregated. One approach may be using a seed polygon (i.e. a polygon slightly larger than the separability limit, with the other polygons being smaller than the limit) and merge the other small polygons within the limit radius of the seed.

Area distribution: The areas gained from eliminating polygons should be evenly distributed based on the weight of geological units (to avoid the effect of "the rich get richer, the poor get poorer").

The next section introduces some first experiments using some of the above constraints.

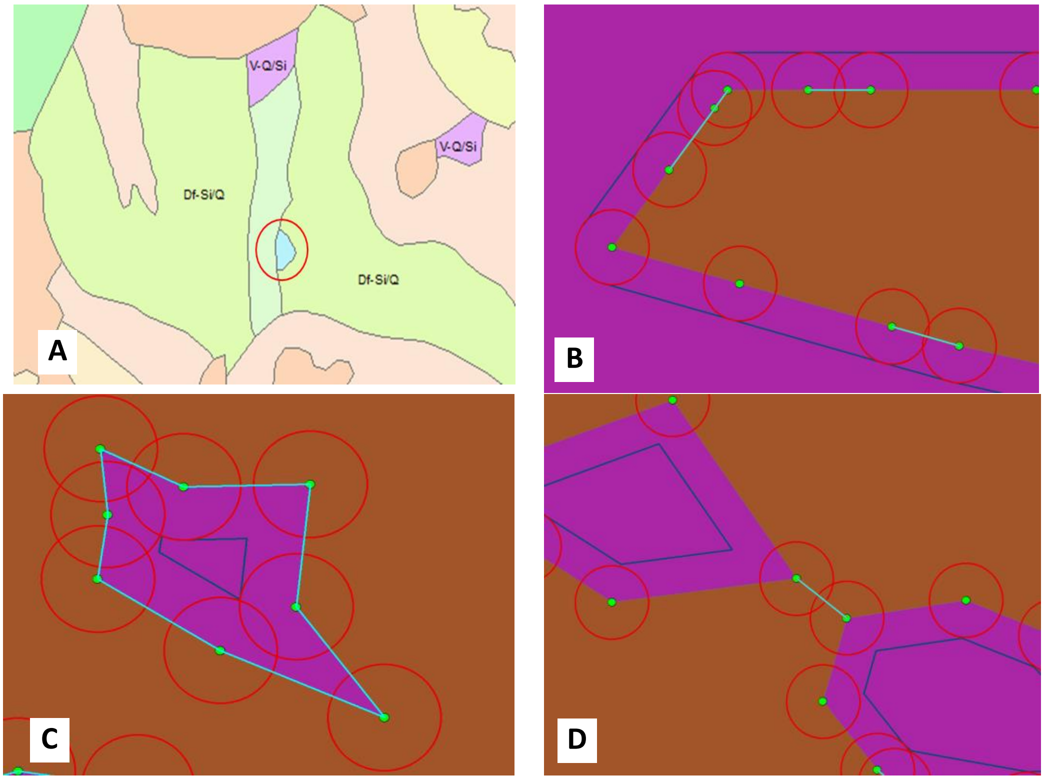

Figure 2. *Describing constraints based on the map content.*

6. Experiments and result

For the experiment a geological map of scale 1:25 000 (Stevens B.P.J et al. 1997) was used as input data. ESRI ArcGIS 10.2, Python 2.7, and morphological filters from OpenCV (Open Source Computer Vision Library) were implemented in different combinations.

The first step of the experiment aimed to identify polygons that are smaller than the minimum size unit and then decide whether to eliminate or to retain and enlarge those polygons. The regular selection tool identified 143 polygon areas with less than the minimum separability unit from 2963 polygons. In order to eliminate polygons violating the constraint two basic principles were used. First, merging polygons with neighboring polygons, which have the largest area, or second, merging polygons with the neighboring polygons, which have the longest shared border (Fig. 3).

The second step was identifying the minimum distance between any part of the polygon and between polygons. This was also carried out by two approaches. The first uses the point distance tool in ArcGIS, which calculates the distance between points; subsequently the places violating the constraints are identified (Fig. 4). The second approach applies buffering in order to find conflict among polygons (as intersecting buffers). This approach can be used between polygons as well as to check whether the minimum interior width is maintained.

One of the solutions for this case could be using raster generalization, such as the morphological filters ‘Open’, ‘Close’, ‘Erode’ and ‘Dilate’, which in a sense supports the concept ‘The rich get richer and the poor get poorer’. Figure 5 shows an experiment using the ‘Dilate’ filter, which implements the concept ‘the rich get richer’ and may improve the polygons to be more readable. This filter is controllable using different convolution sizes, and also generalizes tree-like shapes and makes small polygons being merged with bigger neighboring polygons.

Conversely, the ‘Erosion’ filter works the other way round and assists the concept ‘the poor get poorer’. It is also controlled by means of the kernel size by shrinking pixels in the foreground, though increasing pixel inside the areas.

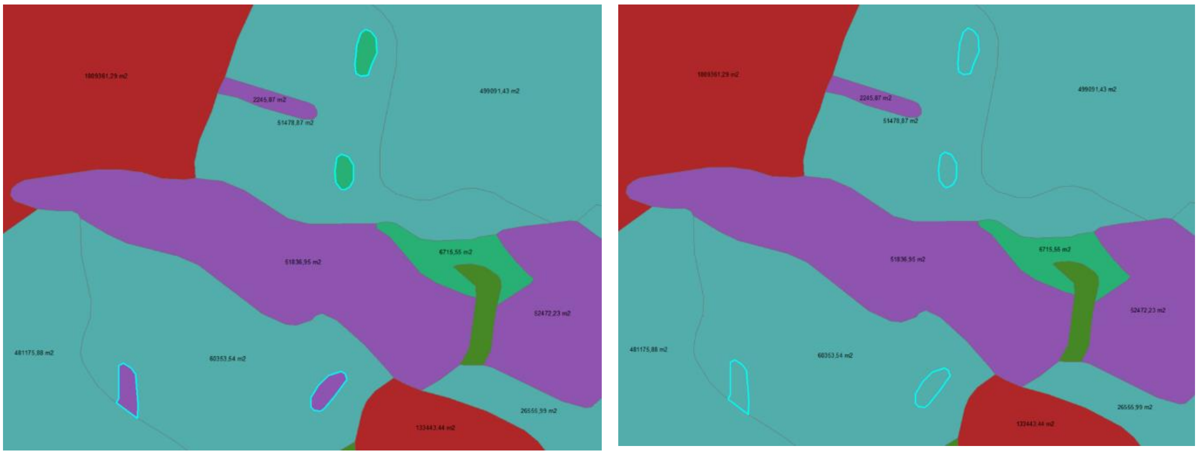

Figure 3. *Identifying and merging polygons that are too small with a neighbouring polygon. Left: Before merging. Right: After merging polygons.*

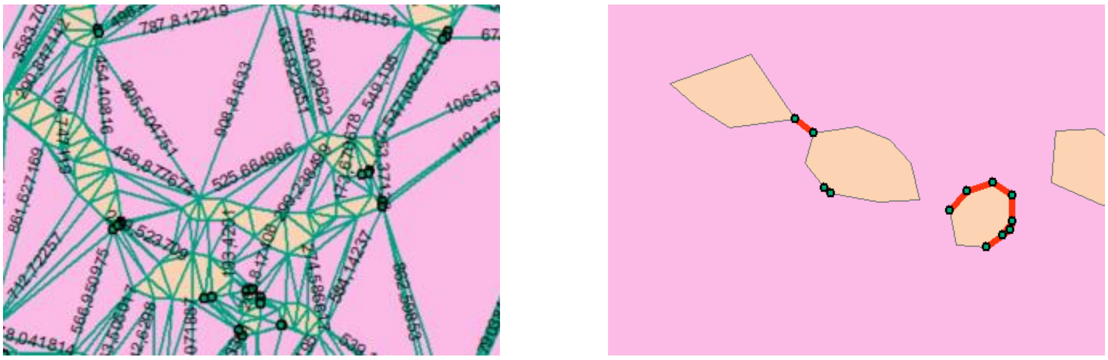

Figure 4. *Left: Identification of minimum distance between points. Right: Finding places that do not meet the constraint of minimum distance between vertices.*

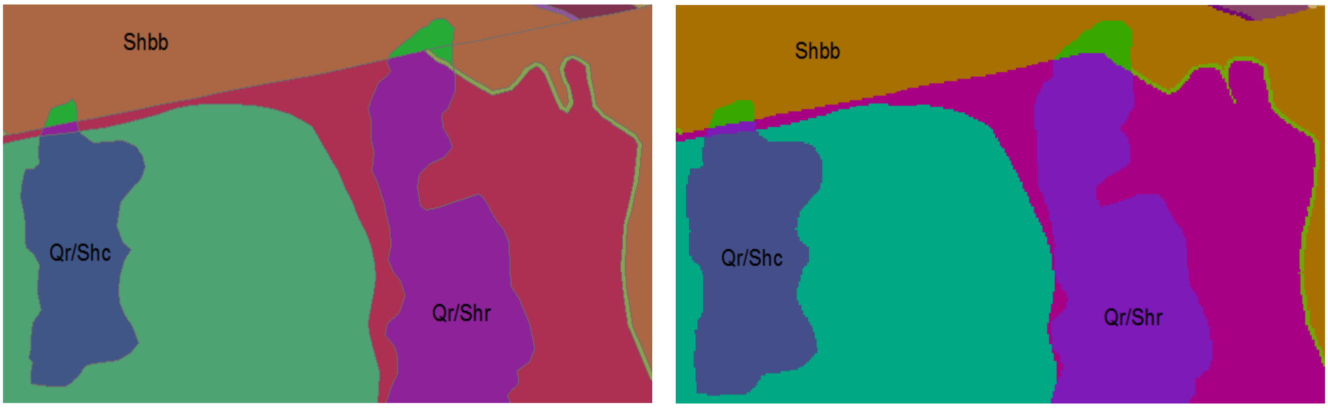

Figure 5. *Application of 'Dilate' filter in order to remove small polygons and obtain a generalized view of the shapes. Left: The original map. Right: The filtered map using Dilate 3x3 kernel size.*

7. Conclusions

Automating the generalization of geological maps can be made more objective and flexible by integrating vector and raster generalization techniques and guiding and monitoring the process with predefined constraints. Defining constraints, taking into account the properties and peculiarities of geological maps, however, is a key point accompanied by logical and structural integration of generalization algorithms. It does not only require generalization algorithms, but also algorithms that implement the measures needed to assess whether the constraints are maintained. Our simple first experiments have shown the general approach, but further research must develop and test algorithms to eventually obtain a complete toolbox of measurement and generalization algorithms.

Acknowledgments

This paper represents part of the PhD project of the first author. Funding by the Swiss Federal Commission for Scholarships (FCS) is gratefully acknowledged.

References

Beard, K., 1991, Constraints on Rule Formation, in Map Generalization: Making Rules for Knowledge Representation (eds Buttenfield, B. P. and McMaster, R. B.), Longman, pp.121-135.

Brassel, K.E. & Weibel, R. (1988). A Review and Framework of Automated Map Gener-alization. Int. Journal of Geographical Information Systems (IJGIS), 2(3): 229-244.

Downs, T. C., & Mackaness, W. a. (2002). An Integrated Approach to the Generalization of Geological Maps. *Cartographic Journal, The*, *39*(2), 137–152. http://doi.org/10.1179/000870402786962489.

Galanda, M. (2003). Automated Polygon Generalization in a Multi Agent System. PhD thesis, Department of Geography, University of Zurich, 188 pages. Retrieved from http://www.geo.uzh.ch/fileadmin/files/content/abteilungen/gis/research/phd_theses/thesis_MartinGalanda_2003.pdf

Harrie, L., & Weibel, R. (2007). Chapter 4 Modelling the Overall Process of Generalisation. *Generalisation of Geographic Information: Cartographic Modelling and Applications*, 1–21. http://doi.org/http://dx.doi.org/10.1016/B978-008045374-3/50006-5.

Jirsa Mark A. and Boerboom Terrence J. COMPONENTS OF GEOLOGIC MAPS 2003. http://docslide.us/documents/components-of-geologic-maps-by-mark-a-jirsa-and-terrence-j-boerboom-2003.html.

Lepland, A. (2014). Marin Geological Interpretations Digitizing scale, details and methods, 1–11.

Li, Z., & Openshaw, S. (1992). Algorithms for automated line generalization 1 based on a natural principle of objective generalization. *International Journal of Geographical Information Systems*, *6*(5), 373–389. http://doi.org/10.1080/02693799208901921.

Li, Z., & Openshaw, S. (1993). A Natural Principle for the Objective Generalization of Digital Maps. *Cartography and Geographic Information Science*, *20*(1), 19–29. http://doi.org/10.1559/152304093782616779.

Marjoribanks, R. (2010). Geological methods in mineral exploration and mining. Geological Methods in Mineral Exploration and Mining. http://doi.org/10.1007/978-3-540-74375-0

McCabe, C., 2008. Vector Approaches to Generalizing Faults and Polygons in 1:24,000 Geologic Maps: Santa Rosa, California, Case Study. http://www.geovista.psu.edu/publications/2008/CMcCabe_GeologicMapGeneralization.pdf.

Peter, B., & Weibel, R. (1999). Using vector and raster-based techniques in categorical map generalization. Third ICA Workshop on Progress in Automated Map Generalization, 14 pgs.

Ruas, A. (1998). OO-Constraint Modeling to Automate Urban Generalization Proceeds, Proceedings of the 8th International Symposium on Spatial Data Handling, Vancouver, BC, 6b, pp. 225-235.

Ruas, A., Plazanet, C. (1996). Strategies for Automated Generalization. Proceedings of the 7th Intl. Symposium on Spatial Data Handling, Deft, The Netherlands, pp. 6.1- 6.18.

Schylberg, L. (1993). Computational Methods for Generalization of Cartographic Data in a Raster Environment. Doctoral Thesis, Royal Institute of Technology, Stockholm, Sweden.

Shea,K.S., and McMaster,R.B (1989). Cartographic Generalization in a Digital Environment When and How to Generalize, AUTO-CARTO 9, Baltimore, Maryland, pp56-65

Smirnoff, A., Huot-Vézina, G., Paradis, S. J., & Boivin, R. (2012). Generalizing geological maps with the GeoScaler software: The case study approach. *Computers and Geosciences*, *40*, 66–86. http://doi.org/10.1016/j.cageo.2011.07.013.

Smirnoff, A., Paradis, S. J., & Boivin, R. (2008). Generalizing surficial geological maps for scale change: ArcGIS tools vs. cellular automata model. *Computers and Geosciences*, *34*(11), 1550–1568. http://doi.org/10.1016/j.cageo.2007.10.013.

Spiess, E., Baumgartner, U., Arn, S., & Vez, C. (2002). Topographic Maps: Map Graphics and Generalisation. Cartographic Publication Series Vol. 17, Swiss Society of Cartography.

Steiniger, S. & Weibel, R. (2005). A Conceptual Framework For Automated Generalization and its Application to Geologic and Soil Maps. *Proceedings 22nd International Cartographic Conference*, La Coruña (Spain), 11-16 July 2005.

Stevens B.P.J, Willis I.L, and Brown R.E 1997 Euriowie-Campbells Creek 1:25,000 Geological Sheet, 7235-III-S and 7135-II-S. New South Wales Geological Survey, Sydney.

Stoter, J., Baella, B., Blok, C., Burghardt, D., Davila, F., Duchene, C., Pla, M., Regnault, N., Touya, G., (2010). EuroSDR research on state-of-the-art of automated generalization in commercial software: Main findings and conclusions. In: Proceedings of the 13th ICA Workshop on Generalization and Multiple Representation, 13–14 September, Zurich, Switzerland, 6 pp.

Weibel, R., Dutton, G.H. (1998). Constraints-Based Automated Map Generalization. Proceedings of the 8th International Symposium on Spatial Data Handling, Vancouver, BC, pp.214-24.